Journal of Biomedical Informatics 51 (2014) 3–7

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Commentary

Technical desiderata for the integration of genomic data with clinical decision support

Brandon M. Welch [a,b,*], Karen Eilbeck [a,c], Guilherme Del Fiol [a], Laurence J. Meyer [d,e], Kensaku Kawamoto [a]

[a] Department of Biomedical Informatics, University of Utah, Salt Lake City, UT, United States
[b] Program in Personalized Health Care, University of Utah, Salt Lake City, UT, United States
[c] Department of Human Genetics, University of Utah, Salt Lake City, UT, United States
[d] Departments of Dermatology and Internal Medicine, University of Utah, Salt Lake City, UT, United States
[e] Department of Pediatrics, University of Utah, Salt Lake City, UT, United States

ARTICLE INFO

Article history:
Received 21 October 2013
Accepted 29 May 2014
Available online 12 June 2014

Keywords:
Clinical decision support systems
Medical genetics
Genomics
Electronic health records
Health information technology
Personalized medicine

ABSTRACT

The ease with which whole genome sequence (WGS) information can be obtained is rapidly approaching the point where it can become useful for routine clinical care. However, significant barriers will inhibit widespread adoption unless clinicians are able to effectively integrate this information into patient care and decision-making. Electronic health records (EHR) and clinical decision support (CDS) systems may play a critical role in this integration. A previously published technical desiderata focused primarily on the integration of genomic data into the EHR. This manuscript extends the previous desiderata by specifically addressing needs related to the integration of genomic information with CDS. The objective of this study is to develop and validate a guiding set of technical desiderata for supporting the clinical use of WGS through CDS. A panel of domain experts in genomics and CDS developed a proposed set of seven additional requirements. These desiderata were reviewed by 63 experts in genomics and CDS through an online survey and refined based on the experts' comments. These additional desiderata provide important guiding principles for the technical development of CDS capabilities for the clinical use of WGS information.

1. Introduction

Rapid genomic sequencing, including whole genome sequencing (WGS) and exome sequencing, is the future paradigm of clinical genetic testing [1]. With a patient's entire genome readily available to a clinician at the point of care, WGS may offer many benefits to traditional single gene testing. Currently, the effective use of single gene testing is inhibited by several factors including the need for clinical indication prior to ordering, the time delay between test ordering and the return of results, and financial constraints for single gene tests [2]. WGS will likely overcome many such barriers in the future. For instance, with WGS information readily available, pathogenic variants in disease-causing genes can be made known to the clinician much earlier in the decision-making process to aid in differential diagnosis [3]. Likewise, important but clinically under-utilized use cases such as pharmacogenomics may now become more clinically and financially feasible than under the current genetic testing paradigm [4]. Indeed, the cost and value of WGS is now at a point at which health care providers are sequencing a patient's genome for unique clinical scenarios [5,6]. As a result, it may be soon when WGS becomes a larger part of routine clinical care [1].

1.1. Challenge of genome data in the clinical setting

While WGS offers many opportunities to enhance clinical care, were it to be made widely available for routine clinical care today, the effective use of WGS information would be hindered by significant barriers. These barriers include inadequate laboratory reporting methods, the complexity of genetic analysis, lack of physician proficiency in genetic analysis, and the insufficient number of genetics professionals in the workforce [7,8]. As clinicians are already burdened with significant time constraints, adding an additional layer of WGS information that they are required to review, integrate with other clinical parameters, and translate into appropriate clinical actions is unlikely to be successful without assistance [9]. This challenge is amplified by the rapidly evolving

* Corresponding author. Address: Biomedical Informatics Center, Department of Public Health Informatics, 135 Cannon Street, Suite 405L, MSC 200, Charleston, SC 29425, United States.
E-mail address: welchbm@musc.edu (B.M. Welch).

http://dx.doi.org/10.1016/j.jbi.2014.05.014
1532-0464/

Table 1
Potential examples of CDS leveraging WGS data.

CDS type	Clinical genomics example
Medication dosing support	CDS automatically adjusts warfarin dosing as a result of known alleles in the VKORC1 and CYP2C9 genes
Order facilitators	An order for colonoscopy is recommended at a younger age as a result of known pathogenic mutations in genes associated with colon cancer
Alerts and reminders	During medication ordering, gene variants known to affect drug pharmacokinetics are checked and clinicians are alerted to potential gene-drug interactions
Relevant information display	Context aware infobuttons in the problem list leverage genome data to provide genetic risk information for a patient with breast cancer
Expert systems	The EHR provides a 10-year cardiovascular disease risk score based on clinical, environmental, and genetic risk factors
Workflow support	The EHR schedules a genetic counseling consultation during prenatal visit due to presence of an X-linked disease gene variant

Table 2
Desiderata for the integration of genomic data into EHRs described by Masys et. al.

1	Maintain separation of primary molecular observations from the clinical interpretations of those data
2	Support lossless data compression from primary molecular observations to clinically manageable subsets
3	Maintain linkage of molecular observations to the laboratory methods used to generate them
4	Support compact representation of clinically actionable subsets for optimal performance
5	Simultaneously support human-viewable formats and machine-readable formats in order to facilitate implementation of decision support rules
6	Anticipate fundamental changes in the understanding of human molecular variation
7	Support both individual clinical care and discovery science

Table 3
Additional desiderata for the technical integration of WGS with CDS.

8	CDS knowledge must have the potential to incorporate multiple genes and clinical information
9	Keep CDS knowledge separate from variant classification
10	CDS knowledge must have the capacity to support multiple EHR platforms with various data representations with minimal modification
11	Support a large number of gene variants while simplifying the CDS knowledge to the extent possible
12	Leverage current and developing CDS and genomics standards
13	Support a CDS knowledge base deployed at and developed by multiple independent organizations
14	Access and transmit only the genomic information necessary for CDS

nature of our understanding of the genome and its clinical implications [10].

1.2. Potential of clinical decision support

Clinical decision support (CDS) integrated into the clinician's workflow provides a practical solution to allow clinicians to provide effective clinical care using genomic information [11,12]. CDS entails providing clinicians, patients, and other healthcare stakeholders with pertinent knowledge and person-specific information, intelligently filtered or presented at appropriate times, to enhance health and healthcare [13]. Examples of point-of-care CDS include medication dosing support, order facilitators, alerts and reminders, relevant information display, expert systems, and workflow support [14]. CDS knowledge, which supports these types of CDS, is used to process clinical data to provide patient-specific advice, recommendations, or information. Table 1 provides summary of CDS types and potential examples of WGS-enabled CDS.

When developed and implemented properly, CDS has the ability to process large amounts of complex data, such as WGS data, and present actionable, evidence-based recommendations to clinicians at the point of care [15]. In doing so, CDS has been shown to be effective in reducing errors, improving clinician performance, and ultimately improving the quality of care in clinical settings [16]. Furthermore, CDS is able to translate research discoveries into clinical care much more efficiently than other traditional methods of knowledge translation [17]. Indeed, CDS may be essential to meeting the demands of WGS at the point of care [18].

1.3. Masys et al. desiderata for integration of genomic information with EHRs

To provide guidance on how to integrate genomic information within the EHR, Masys et al. [19] developed a set of guiding principles for the technical integration of genomic information into the EHR. Their summarized desiderata is outlined in Table 2. While this paper provides a strong foundation for integration of genomic information within the EHR, in our experience developing CDS capabilities for WGS, we found that it did not fully address all the needs of WGS integration with CDS.

To address this need, we assembled a core group of domain experts in genomics and CDS to define additional desirable functional characteristics for CDS capable of incorporating WGS at the point of care within an EHR. These additional requirements were then validated and assessed for importance among a larger group of domain experts in genomics and CDS. These additional desiderata, described in Table 3, are intended to augment the original Masys et al. desiderata and provide further guidance to system developers on important requirements to consider when developing health IT systems and CDS for WGS information. Of note, the requirements described in this desiderata represent current and potential needs according to our current understanding. Nevertheless, future research and development may require additional requirements to be added or compel current requirements to be removed from the desiderata. As such, this initial set of requirements should be viewed as an evolving set of requirements that informs and is informed by ongoing research and development in this field.

2. Development of additional desiderata

A core group of domain experts in genomics and CDS was assembled to review needs for WGS CDS. This group of domain experts derived a set of additional desiderata by consensus. The members of the core group consisted of the authors of this manuscript and described in Appendix A. We sought to improve, refine, and validate our initial set of requirements by seeking additional input and qualitative feedback from a wider community of domain experts in genomics and CDS. To obtain this feedback, an anonymous, IRB-approved survey was distributed electronically to e-mail discussion lists of relevant expert groups [20]. Participants included members of the HL7 Clinical Genomics Workgroup, HL7 CDS Workgroup, AMIA Genomics Workgroup, AMIA CDS Workgroup, Open Source Electronic Health Record Agent (OSEHRA) Genomics Workgroup, the developers of ClinVar, University of Utah Program in Personalized Health Care, and our own professional contacts. The survey instrument was available from 8/13/2013 through 8/30/2013. Each requirement was listed and summarized on a separate page with a 5-point Likert scale assessing the participant's opinion regarding the importance of each requirement (very important, important, neither important nor unimportant, unimportant, very unimportant, and unsure/no opinion/blank). The full survey is available in Appendix B. Survey responses were analyzed and visualized using spreadsheet software. Qualitative feedback was reviewed by the core panel, and feedback was incorporated into the desiderata where appropriate. Participant feedback, resulting modifications, and core group responses are available in Appendix C.

3. Additional desiderata for the technical integration of WGS with CDS

The final desiderata is summarized in Table 3, listed in order of importance as assessed by the community of experts. All proposed requirements were judged to be important (with high significance) by the community of experts. The explanation and reasoning for each additional requirement is provided here.

3.1. CDS knowledge must have the potential to incorporate multiple genes and clinical information. (Requirement #8)

A relatively small number of Mendelian diseases, such as cystic fibrosis and sickle-cell anemia, are affected by variants within a single gene responsible for producing the characteristic phenotype. As a result, such cases are fairly straightforward to assess. With nearly every human condition affected one way or another by a genetic influence, most diseases, in particular common diseases, are caused or affected by multiple genetic influences and environmental factors. For example, there are potentially hundreds of genetic loci contributing to type 2 diabetes risk [21]. In order to provide an accurate risk assessment and decision support, all relevant genetic loci, relevant clinical factors (e.g. age, weight, health history, co-morbidities), and environmental influences (e.g. diet, physical activity, stress) need to be considered. Furthermore, it may not always be the case that all necessary data is in one central location. Therefore, CDS for the WGS must have the capacity to leverage and incorporate several pieces of information from multiple genomic and non-genomic data sources.

3.2. Keep CDS knowledge separate from the variant classification. (Requirement #9)

Masys et al. describe the importance of separating molecular observations (e.g. DNA sequence) from variant classification (e.g. pathogenicity classifications) due to the need to update variant interpretation as genome knowledge changes and grows over time. To illustrate, one study found that over a seven year period, 14.5% of reported variant classifications had to be reclassified [22]. Likewise, it is also essential to separate CDS knowledge from both molecular observations and variant classification. CDS must have the ability to manage evolving and frequently changing gene variant interpretations efficiently without requiring changes to the underlying CDS knowledge each time a variant's classification changes. Separation of CDS knowledge from variant interpretation allows CDS knowledge to be more efficiently handled and maintained.

3.3. Have the capacity to support multiple EHR platforms with various data representations with minimal modification. (Requirement #10)

The reality of the health information environment in the US today is that multiple healthcare organizations use multiple EHR and health information management systems [23]. Often, these health information management solutions store and represent the same health information differently. This can be a challenge when trying to harness the information within different health IT systems in different organizations to provide CDS. Due to the need to distribute and share WGS enabled CDS knowledge across multiple organizations (see next requirement), the CDS architecture would ideally be EHR agnostic, where CDS knowledge can be developed once, and then run consistently anywhere. A number of initiatives aimed at supporting this type of architecture are underway, including the Health eDecisions initiative, OpenCDS, the SMART platform, and the CDS Consortium, to name a few.

3.4. Support a large number of gene variants while simplifying the CDS knowledge to the extent possible. (Requirement #11)

There are roughly 1200 known variants in the adenomatous polyposis coli (APC) gene, a gene associated with a rare form of colon cancer [24]. Likewise, there are nearly 2000 known variants in the cystic fibrosis gene CFTR [25]. Given the potentially high number of variants per gene, it may be inefficient to create CDS knowledge for every known variant in each disease-causing gene. Furthermore, as novel variants are discovered, it will be difficult to update CDS knowledge for every gene variant that is discovered. Therefore, to manage this complexity, variants with the same or similar clinical impact should be classified accordingly. CDS knowledge can then be simplified by developing logic or rules which leverages the variant interpretation rather than the specific variant. Nevertheless, in cases where a particular variant has a unique and clinically important impact or where machine learning CDS models could utilize individual variants or combinations of variants within a single gene, genetic information at the variant level should still be accessible to CDS knowledge. In summary, CDS knowledge can be greatly simplified by classifying variants into groups of common clinical impact, while still supporting inferencing at the individual variant level where necessary.

3.5. Leverage current and developing CDS and genomics standards. (Requirement #12)

Both the CDS and genomics fields have benefited from extensive research and development over the years. Indeed, both fields have well developed infrastructure and standards to support its uses. Therefore, it is important to leverage these standards and infrastructure. Examples include using HGVS and dbSNP to represent specific molecular observations; ACMG recommendations for variant classification; HL7 Clinical Genomics standards for the representation of genetic information; Arden Syntax, GELLO, GLIF3,

GEM, or the HL7 CDS Knowledge Artifact Implementation Guide for CDS knowledge representation; the HL7 Decision Support Service standard and HL7 Infobutton standard for delivering CDS as a service; and open-source, standards-based resources such as OpenCDS. While many of these standards and resources may not be completely sufficient for meeting the needs of CDS for WGS, it still represents significant relevant effort. It will be important to leverage current and developing CDS and genomics infrastructure, standards, and knowledge.

3.6. Support a CDS knowledge base deployed at and developed by multiple independent organizations. (Requirement #13)

With the potential for genomic information to impact nearly every clinical decision and the clinical application of genomics rapidly evolving, the time and cost for a single entity or organization to manually create and update CDS knowledge will be challenging. Indeed, it is unlikely that a single health care organization will be able to author and manage all CDS knowledge for all WGS use cases. Furthermore, there should be an efficient and scalable mechanism to consistently modify CDS knowledge everywhere it is deployed. Ideally, a standardized CDS infrastructure would allow multiple health care organizations, public or private entities, or individuals to create, publish, and distribute CDS knowledge efficiently to multiple consuming health care organizations. Such an approach will allow a specialized entity (e.g. pharmacogenomics experts) to develop and manage CDS knowledge and subsequently distribute to 'subscribing' health care organizations. With an ecosystem of CDS knowledge developed independently by multiple content developers, it becomes more feasible for health care organizations to have affordable access to comprehensive, up-to-date, and accurate CDS knowledge for the entire genome.

3.7. Access and transmit only the genomic information necessary for CDS. (Requirement #14)

The separation of CDS knowledge from molecular observations and variant interpretations will require relevant genetic information being accessed and sent to a CDS engine (or equivalent) for processing. It will be inefficient and insecure to transmit an entire genome file for every CDS knowledge. The processing capacity required to transmit and sift through an entire genome for CDS knowledge will hinder the ability to provide CDS at the point of care in real-time. Furthermore, HIPAA requires that only the minimum protected health information needed to satisfy a particular purpose or carry out a function be used or transmitted [26]. Therefore, a CDS architecture must only transmit the relevant genes and any associated molecular observations, variant interpretations, and associated clinical information.

4. Discussion

To guide development of CDS capabilities for WGS information, we identified several additional requirements that were not addressed in a previously published work on integrating genomic information with EHRs [19]. These additional proposed requirements were reviewed by a community of experts and found to be important. While the Masys et al. desiderata primarily focused on the integration of genomic information within the EHR, our desiderata focus largely on the integration of genomic information with CDS capabilities. The combination of both desiderata is important to leverage WGS within EHRs using CDS. Given the important need to provide CDS capable of supporting WGS information at the point of care, and the requirements to support both the Masys et al. and these desiderata, a novel CDS architecture capable of supporting these requirements will need to be designed, developed, and evaluated. Due to the scope and complexity of integrating WGS information with CDS as outlined in the desiderata, we propose a CDS architecture that utilizes principles of a service-oriented architecture (SOA). SOA is a software design methodology based on a collection of separate, independent software components known as services, which are self-contained and have well-defined capabilities [27]. Accordingly, we are currently developing a prototype CDS architecture that is based on the principles of SOA and is capable of supporting WGS information in a manner consistent with the desiderata described in this manuscript and the Masys' paper. Once such a prototype is developed and tested, clinical scenarios will test the feasibility of this approach. Finally, once testing is complete, implementing as a pilot study will be important to validate the solution in a real clinical setting. As a result of these efforts, it is likely that additional requirements may be identified for inclusion in the desiderata. As such, this proposed set of desiderata should not be considered a final authoritative set, but rather a foundation upon which further requirements may be added in the future.

Competing Interests

KK is currently or recently served as a consultant on CDS to the Office of the National Coordinator for Health IT, ARUP Laboratories, McKesson InterQual, ESAC, Inc., Inflexxion, Inc., Intelligent Automation, Inc., Partners HealthCare, and the RAND Corporation. KK receives royalties for a Duke University-owned CDS technology for infectious disease management known as CustomID that he helped develop. KK was formerly a consultant for Religent, Inc. and a co-owner and consultant for Clinica Software, Inc., both of which provide commercial CDS services, including through use of a CDS technology known as SEBASTIAN that KK developed. KK no longer has a financial relationship with either Religent or Clinica Software. BMW, GDF, LJM, and KE have no conflicts of interest to disclose.

Funding

This study was funded by the University of Utah Program in Personalized Health Care and Department of Biomedical Informatics. REDCap was available for use in this study under the Grant No. 8UL1TR000105 (formerly UL1RR025764) from National Center for Advancing Translational Sciences and the National Institutes of Health. BMW was supported by the Program in Personalized Health Care. KE was supported by Grant No. 5R01HG004341 from the National Human Genome Research Institute. GDF was supported by Grant No. K01HS018352 from the Agency for Healthcare Research and Quality. The funding sources played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the manuscript; or in the decision to submit the manuscript for publication. The views expressed in this manuscript are those of the authors alone and do not necessarily reflect the official views of the University of Utah, the National Center for Advancing Translational Sciences, the National Institutes of Health, the National Human Genome Research Institute, or the Agency for Healthcare Research and Quality.

Contributor

All authors contributed to the design and conducted of the study, as well as the preparation of the manuscript. BMW implemented the study and performed the analysis.

Acknowledgments

The authors would like to thank the HL7 Clinical Genomics Workgroup, HL7 CDS Workgroup, AMIA Genomics Workgroup, AMIA CDS Workgroup, Open Source Electronic Health Record Agent (OSEHRA) Genomics Workgroup, and the University of Utah Program in Personalized Health Care for allowing us to distribute our survey to their members.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2014.05.014.

References

[1] Bonetta L. Whole-genome sequencing breaks the cost barrier. Cell 2010:917–9.
[2] Weldon CB, Trosman JR, Gradishar WJ, Benson AB, Schink JC. Barriers to the use of personalized medicine in breast cancer. J Oncol Pract 2012;8:e24–31.
[3] Feero WG, Guttmacher AE, Collins FS. Genomic medicine – an updated primer. N Engl J Med 2010;362:2001–11.
[4] McWilliam A, Lutter R, Nardinelli C. Health care savings from personalizing medicine using genetic testing: the case of warfarin. AEI-Brookings Joint Center for Regulatory Studies; 2006.
[5] Saunders CJ, Miller NA, Soden SE, Dinwiddie DL, Noll A, Alnadi NA, et al. Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Sci Transl Med 2012:4. 154ra135.
[6] Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, et al. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. Science 2010;328:636–9.
[7] Green RC, Rehm HL, Kohane IS. Clinical Genome Sequencing, 2nd ed. Genomic Pers. Med.; 2013, p. 102–22.
[8] Schrijver I, Aziz N, Farkas DH, Furtado M, Gonzalez AF, Greiner TC, et al. Opportunities and challenges associated with clinical diagnostic genome sequencing: a report of the association for molecular pathology. J Mol Diagnostics 2012;14:525–40.
[9] Masys DR. Effects of current and future information technologies on the health care workforce. Health Aff 2002;21:33–41.
[10] Shirts BH, Parker LS. Changing interpretations, stable genes: responsibilities of patients, professionals, and policy makers in the clinical interpretation of complex genetic information. Genet Med 2008;10:778–83.
[11] Welch BM, Kawamoto K. Clinical decision support for genetically guided personalized medicine: a systematic review. J Am Med Inform Assoc 2012;20:388–400.
[12] Kawamoto K, Lobach DF, Willard HF, Ginsburg GS. A national clinical decision support infrastructure to enable the widespread and consistent practice of genomic and personalized medicine. BMC Med Inform Decis Mak 2009;9:17.
[13] Osheroff JA, Teich JM, Middleton B, Steen EB, Wright A, Detmer DE. A roadmap for national action on clinical decision support. J Am Med Informatics Assoc JAMIA 2007;14:141–5.
[14] Wright A, Sittig DF, Ash JS, Feblowitz J, Meltzer S, McMullen C, et al. Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems. J Am Med Inform Assoc 2011;18:232–42.
[15] Bennett C, Doub T, Selove R. EHRs connect research and practice. Where predictive modeling, artificial intelligence, and clinical decision support intersect. Heal Policy Technol 2012:1–24.
[16] Jaspers MWM, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011;18:327–34.
[17] Sarkar IN. Biomedical informatics and translational medicine. J Transl Med 2010;8:22.
[18] Ullman-Cullere MH, Mathew JP. Emerging landscape of genomics in the electronic health record for personalized medicine. Hum Mutat 2011;32:512–6.
[19] Masys DR, Jarvik GP, Abernethy NF, Anderson NR, Papanicolaou GJ, Paltoo DN, et al. Technical desiderata for the integration of genomic data into Electronic Health Records. J Biomed Inform 2012;45:419–22.
[20] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81.
[21] Imamura M, Maeda S. Genetics of type 2 diabetes: the GWAS era and future perspectives. Endocr J 2011;58:723–39.
[22] Aronson SJ, Clark EH, Varugheese M, Baxter S, Babb LJ, Rehm HL. Communicating new knowledge on previously reported genetic variants. Genet Med 2012;14:713–9.
[23] Zhang NJ, Seblega B, Wan T, Unruh L, Agiro A, Miao L. Health information technology adoption in U.S. Acute care hospitals. J Med Syst 2013;37:9907.
[24] APC homepage – Colon cancer gene variant databases – Leiden Open Variation Database n.d.
[25] Cystic Fibrosis Mutation Database: Statistics n.d.
[26] US Department of Health and Human Services Office of the Secretary. Standards for Privacy of Individually Identifiable Health Information; Final Rule; 2000.
[27] Juneja G, Dournaee B, Natoli J, Birkel S. Improving performance of healthcare systems with service oriented architecture. InfoQ 2008:1–15.